Food Research International 130 (2020) 108936

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Compositional identification and authentication of Chinese honeys by ^{1}H NMR combined with multivariate analysis

Chenxi He[a], Yun Liu[b], Huili Liu[c], Xin Zheng[a], Guiping Shen[a], Jianghua Feng[a,*]

[a] Department of Electronic Science, Fujian Provincial Key Laboratory of Plasma and Magnetic Resonance, Xiamen University, Xiamen 361005, China
[b] Nanjing Customs, Animal Plant and Food Inspection Centers, Nanjing 210001, China
[c] State Key Laboratory of Magnetic Resonance and Atomic and Molecular Physics, Wuhan Center for Magnetic Resonance, Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Wuhan 430071, China

ARTICLE INFO

Keywords:
High resolution nuclear magnetic resonance
Food quality
Multivariate statistical analysis
Authenticity
Honey

ABSTRACT

Honey authentication has been becoming more and more important and necessary to the honey producers, the consumers and the market regulatory authority due to its favorite organoleptic and healthy properties, high value and increasing export but prevalent falsification practice for economic motivation in China and the potential health risk of adulterated honey. In this study, we obtained the spectral profiles of 90 authentic and 75 adulterated Chinese honey samples by means of high resolution nuclear magnetic resonance (NMR) spectroscopy, and 65 kinds of major and minor components in honey were identified and quantified from their NMR spectra. Combining with the multivariate statistical analyses including principal component analysis (PCA), linear discriminant analysis (LDA), and orthogonal partial least squared-discriminant analysis (OPLS-DA), the discrimination models were successfully established to identify the adulterated honeys from the authentic ones with an accurate rate of 97.6%. Furthermore, the corresponding volcano plot was used to screen out 8 components including proline, xylobiose, uridine, β-glucose, melezitose, turanose, lysine and an unknown component, which are responsible for the differentiation between the authentic and adulterated honeys and will help to control Chinese domestic honey market.

1. Introduction

Honey is a natural sweet substance produced by honey bees and usually used as a sweetener because of its positive nutritional and health effects since antiquity (Spiteri et al., 2017). The main components of honey are carbohydrates (70–80% w/w) composed by glucose (~31% w/w) and fructose (~38% w/w) and water (10–20% w/w) (Ouchemoukh, Louaileche, & Schweitzer, 2007), and a great number of minor components also exist in honey. The minor components mainly include organic acids, amino acids, poly-saccharides, proteins, enzymes, lipids, vitamins, volatile chemicals, phenolic acids and minerals (Amiry, Esmaiili, & Alizadeh, 2017).

According to Codex Alimentarius standard, authentic honey should be a natural foodstuff that prohibits from adding any food ingredient. However, thanks to the healthy property and unique flavor character of honey, economic motivation makes this high-value foodstuff particular attractive by intentional addition of cheaper syrups (Wu et al., 2017). It is very challenging to detect sugar-based adulterated honey by general method because of the similarity of chemical compositions between sugar-based adulterants and authentic honeys. It is reported that honey is highly vulnerable to food fraud as it accounts for approximately 90% of all adulterations related to sweeteners in Europe in recent years (Sobrino-Gregorio, Vargas, Chiralt, & Escriche, 2017). China is both a large honey producer and a large worldwide honey provider. It is estimated that annual production of honey exceeds 200,000 tons in China and about half of the outputs are exported to the world market (Li, Shan, Zhu, Zhang, & Ling, 2012). Honey adulteration not only reduces quality of the honey but also downgrades market credit rate in the region. In addition, honey adulterants even may pose a health risk due to the increasing levels of low density lipoprotein and cholesterol (Liu, Qu, Luo, Xu, & Zhong, 2019). Therefore, it is essential to detect and identify adulterated honey for governments, food industry and scientific field.

There is a growing interest to develop analytical techniques to identify the honey adulteration. Thin-layer chromatography is the oldest method for honey analysis and usually used to detect high fructose corn syrup in honey, but it is limited because of false positives caused by hydrolysis of oligosaccharides and polysaccharides (Kushnir, 1979). C-isotope approach is feasible to identify the C3 sugars-based

* Corresponding author.
E-mail address: jianghua.feng@xmu.edu.cn (J. Feng).

https://doi.org/10.1016/j.foodres.2019.108936
Received 8 October 2019; Received in revised form 19 December 2019; Accepted 19 December 2019
Available online 23 December 2019
0963-9969/

adulterated honey but feeble for C4 sugars adulterants (Rogers et al., 2014). Chromatographic methods including gas chromatography (Ruiz-Matute, Rodríguez-Sánchez, Sanz, & Martínez-Castro, 2010), high-performance anion-exchange chromatography (Morales, Corzo, & Sanz, 2008), and high-performance liquid chromatography (Wang et al., 2015) are also widely used to analyze composition of honey with relatively high resolution and sensitivity, but these methods are time-consuming, destructive, and expensive.

Nuclear magnetic resonance (NMR) can quantify the different compounds and provide structural information of compounds in a mixture via a single NMR experiment with excellent repeatability and reproducibility, and it has become a responsible and promising tool to detect food adulteration, such as edible oil (Fang et al., 2013), milk (Santos, Pereira-Filho, & Colnago, 2016), and prawn (Li, Li, & Zhang, 2018). Moreover, NMR is fast (usually only requires 5 min or less to collect a ^{1}H NMR spectrum), non-destructive, high-throughput and without complex sample pretreatment. A few researchers had applied NMR technology to identify the adulterated honeys (Musharraf, Ambreen Fatima, Siddiqui, Iqbal Choudhary, & Atta-Ur-Rahman. , 2016). A commercial tool, Honey-Profiling Module of the NMR Food-Screener of Bruker Corporation, is also available to tackle honey fraud (Sobolev et al., 2019). Recently, some advanced NMR methods were developed for adulteration discrimination of honey. High-resolution pure shift NMR spectroscopy provided a good demonstration of honey authentication by improving the spectral resolution of the high-content components in honey (Bo et al., 2019). Another successful example is to identify the adulterated honey by overfeeding bee colonies with industrial syrups by CSSF-TOCSY-based NMR approach (Schievano, Sbrizza, Zuccato, Piana, & Tessari, 2019). However, most of these researches focused on some specific components in honey (Del Campo, Zuriarrain, Zuriarrain, & Berregi, 2016), and the identification accuracy is unsatisfactory. In this work, we aimed to display a global nutritional profile of Chinese honey to honey producers and consumers and provide a non-commercial discrimination solution of honey adulteration by comprehensively NMR assignment of almost all of minor and major components in Chinese honey. Meanwhile, multivariate statistical analyses were applied to identify relevant markers for honey authentication.

2. Materials and methods

2.1. Samples collection and preparation

A total of 90 authentic honey samples from different botanic and geographical origins and 75 adulterated honey samples with various kinds of syrup were provided by Animal and Plant and Food Testing Center of Jiangsu Entry-Exit Inspection and Quarantine Bureau, China. The botanic origins of authentic honey include *Lycium barbarum*, *Eriobotrya japonica*, locust, Chinese jujube, linden, *Leonurus sibiricus,* chaste, *Brassica napus,* Sunflower, *Lonicera japonica*, *Flos chrysanthemi indici*, *Schefflera actinophylla*, *Dimocarpus longan*, and so on. And their geographic origins include almost all of honey production provinces of China such as Jiangsu, Henan, Xinjiang, Inner Mongolia, Hubei, Liaoning, Shandong, Sichuan, etc. The adulterated honeys were directly collected from market and verified by testing center, and they were revealed in various adulterated forms including different adulterated proportions of starch syrup, rice syrup, invert syrup, etc. The authentication were confirmed by multiple test methods including mass spectroscopy and chromatography by SGS-CSTC Standard Technical Service Co., Ltd. Diverse and complicated forms of true and adulterated honey were a huge challenge for discrimination.

A total of 165 honey samples were collected into Eppendorf tubes and stored under the condition of 17–25 °C to avoid crystallization (Venir, Spaziani, & Maltini, 2010). Potential of hydrogen (pH) of each honey sample was measured on a SevenCompact pH meter (Mettler Toledo, Switzerland) to decide pH value of the subsequent phosphate buffer. As a result, measured values ranged from 3.0 to 5.5, and therefore the pH value of the buffer solution was set as 4.9.

Honey samples were prepared by dissolving 150 mg of honey in 600 μL of 600 mM of deuterated phosphate buffer (pH 4.9) containing 0.05% of sodium 3-(trimethylsilyl)-2, 2,3,3-$^2H_4$ propionate (TSP). The honey–buffer mixture was put on the table for 5 min at room temperature and then centrifuged at 10,000 rpm for 10 min. Afterwards, the supernatant solution (550 μL) was transferred into a 5-mm NMR tube for NMR measurement.

2.2. ^{1}H NMR spectroscopy and spectral preprocessing

The ^{1}H NMR spectra of all of the honey samples were acquired on a 600 MHz Bruker Ascend™ NMR spectrometer (Bruker Corporation, Karlsruhe, Germany) at 600.38 MHz and 298 K without sample rotation. All the ^{1}H NMR spectra were acquired using water suppressed NOESYPR1D pulse sequence (recycle delay-90°-t_1-90°-t_m-90°-acquisition) with a 9000-Hz spectral width. The acquisition parameters were as follows: 32 scans, time domain 32 K, acquisition time of 2.5 s, delay time of 3.0 s, t_1 of 4 μs, and t_m of 100 ms.

In addition, for accurate quantitation, spin–lattice relaxation time (T_1) of each peak in the NMR spectra was measured using the classical inversion recovery sequence with water suppression during both recycle delay and relaxation delay. Sixteen τ values of 0.05–16 s were employed and 32 K data points of the NMR signal were acquired using 32 scans.

The NMR spectrum of each sample was pre-processed with MestReNova software (V9.0, Mestrelab Reserch, Santiago de Compostela, Spain). All free induction decays (FIDs) were multiplied by an exponential function with a 0.3-Hz line-broadening factor followed by Fourier transformation. The spectra were manually phased and baseline corrected by using the polynomial fit method. Chemical shifts were calibrated by setting the peak of TSP at 0.0 ppm. The regions of water resonance (δ4.81–4.94) were removed to eliminate baseline effects of imperfect water signals. Besides, a so-called segment-wise peak alignment technique implemented by MestReNova was applied to get a proper peak alignment which is critical for the subsequent multivariate statistical analysis. The spectral region of δ0.50–10.00 was automatically integrated with the integral width of 0.005 ppm, and then normalized to the total sum, obtaining 1900 buckets.

A series of standard 2D NMR spectra, including ^{1}H–^{1}H COSY, ^{1}H–^{1}H TOCSY, ^{1}H–^{13}C HSQC, ^{1}H–^{13}C HMBC and J-resolved NMR spectra, were also acquired on the selected honey sample for peak identification purposes.

2.3. Statistical analyses

Multivariate data analysis was performed on SIMCA 14.1 software (Umetrics, Umea, Sweden).

2.3.1. Principle component analysis

Non-targeted principal component analysis (PCA), an unsupervised method, was performed for visualization, dimensionality reduction of the data by linear transformation. In our study, PCA with "mean centering" (Ctr), containing most of the information of the original variable, was used to visualize the preliminary grouping of samples in reduced-dimension data space then classify true honey and adulterated honey.

2.3.2. Linear discriminant analysis

Linear discriminant analysis (LDA) is a supervised method by maximizing the variance between categories and minimizing the variance within categories (Gerhardt et al., 2016). LDA should not use directly raw data since there are multiple collinearities in the variables from ^{1}H NMR data. Therefore, PCA mentioned above was employed to orthogonalize and reduce the dimensionality. Then, PCA score values

were used as the input variables for LDA.

2.3.3. Orthogonal partial least squares discriminant analysis (OPLS-DA)

Non-targeted orthogonal partial least squares discriminant analysis (OPLS-DA), another one supervised method, was an upgrade of partial least squares regression method by removing the systematic orthogonal variation to the response. In our study, four indicators, including R^2 (cumulative), Q^2 (cumulative), $P_{CV\text{-}ANOVA}$ and permutation test results, were used to evaluate model performance. R^2 reveals the goodness of fit, while Q^2 indicates the predictability of the model. The closer the Q^2 value is to 1, the more excellent the model, and values $>$ 0.500 indicate good quality of the model (Schievano, Finotello, Uddin, Mammi, & Piana, 2016). CV-ANOVA (Cross-Validation- Analysis Of Variance) was used to evaluate the significance of OPLS-DA model by calculating a p-value. Permutation test was subsequently used to verify the possible overfitting of the model.

In addition, a four-dimensional enhanced volcano plot was used to identify candidate markers contributing gigantically to the separation of two groups. The fold change was defined as the concentration ratio of a substance in the honey between the two groups, and p-value with Student's *t*-test was analyzed for better reliability of characteristic markers' screening. In our study, $-\log_{10}$(p-value) against $\log_2$(fold-change) was used to represent the y- and x-axes of the volcano plot, respectively (Shen et al., 2018). Therefore, the concentrations for those components located at positive side of horizontal axis in volcano plot are higher in the adulterated honey than in the true honey. The larger the y coordinate, the more significant the marker. Absolute correlation coefficient (|r|) and variable importance projection (VIP) from OPLS-DA model were represented by circle size and color respectively in the scatter plot. The larger circle size represents a larger VIP value, and warm color represents the significant difference between classes, while a cool color is opposite. Generally, candidate markers tend to locate at the upper left or upper right zones of the enhanced volcano plot in larger circle shapes and warmer colors. Typical markers are selected by three criteria: $p < 0.05$, $|r| > 0.500$ and VIP values at the top 10%. The volcano plot was generated via MATLAB scripts (downloaded from http://www.mathworks.com).

2.4. Training and test set selection

To further verify the validity of OPLS-DA and PCA/LDA model. Two thirds of the true and adulterated honey samples were randomly selected to be treated as training set, respectively, the remaining third samples as test set.

3. Results and discussion

3.1. ^{1}H NMR spectral comparison of authentic and adulterated honeys

^{1}H NMR spectrum could provide the global profile of honey and the information of the natural ratios of the different components in honey, and thus spectral comparison between authentic and adulterated honeys could give direct and visual differences in composition. Fig. 1 demonstrated the NMR spectral comparison of authentic and adulterated honeys, and the resonance assignments were performed according to the chemical shift, multiplicity and J coupling constant with the aid of the 2D NMR spectra (COSY, TOCSY, HSQC, HMBC and J-resolved NMR spectra) and related literature data (Del Campo et al., 2016, Consonni & Cagliani, 2008, Boffo, Tavares, Tobias, Ferreira, & Ferreira, 2012), and further confirmed by public Biological Magnetic Resonance Data Bank (BMRB) NMR database. A total of 65 components were identified from the NMR spectra, and the detailed assignment information of the peaks was tabulated in Table 1.

The ^{1}H NMR spectrum of honey could be roughly divided into three regions: aliphatic region (δ0.5–3.0 ppm), carbohydrate region (δ3.0–5.5), and aromatic region (δ5.5–10.0). As shown in Fig. 1, the resonances of the saccharides, especially from glucose and fructose, dominate NMR spectra of honeys, and some other saccharides including mono- (allose, arabinose and rhamnose), di- (erlose, isomaltose, lactose, maltose, sucrose, trehalose, turanose and xylobiose), and tri- (melezitose and raffinose) saccharides are also presented in the NMR spectra. The resonance peaks in the aliphatic region are mainly derived from amino acids, organic acids, organic alcohols and amines. In this region, ethanol signal (δ1.19) is obviously the strongest, followed by rhamnose (δ1.23) signal. In addition, proline also presents considerable concentration. Although there is a large variation of the ethanol signals between the honey samples, it is closely related to storage time and not a typical and reliable marker to differentiate true and adulterated honey (Pontes, Marques, & Câmara, 2007). Regarding the aromatic region, 5-hydroxymethylfurfural (HMF), tyrosine, phenylalanine and formic acid demonstrate higher contents. Some obvious variations in the signal intensity between true and adulterated honeys could be readily recognized by visual inspection, such as higher contents of HMF and pyruvate in the adulterated honey and higher contents of formic acid, phenylalanine, tyrosine, dimethylglycine and methylguanidine in the true honey.

The relative concentrations of any one assigned component could be determined by comparison of the integral area of the corresponding signals with that of the internal standard (TSP) and spin-lattice relaxation time (T_1) correction (Zheng, Zhao, Wu, Dong, & Feng, 2016). Accordingly, the relative concentrations of all kinds of components in the true and adulterated honeys were also listed in Table 1. The concentrations could be considered as the fingerprint of the true honey; however, they need to be compared with those in different geographical and botanical origins and the adulterated honeys in order to identify which chemical components are relevant for sample differentiation. Therefore, further discriminant analysis requires multivariate statistical analysis.

3.2. Discrimination of honey adulteration by PCA/LDA

Principal component analysis (PCA) was performed to reduce the dimensionality of input variables as well as to visualize the differences and similarities between the true and adulterate honeys. PCA score scatter plot demonstrated the metabolically similar cluster of the above 165 honey samples (Fig. 2a). This was probably because the adulterants in these adulterated honey samples were tailored to mimic the natural sucrose-glucose-fructose profile of true honey by the illegal businessmen. However, a part of adulterated honeys are still observed to be obviously separate from the true honeys. This may be due to relatively great compositional difference between these adulterated honey samples and true honey.

According to the loadings line plot in the first two principal components (PCs) (Fig. 2b), the contributions of variables to the separation of different honey samples in the score plot could be revealed, where glucose (including α-glucose and β-glucose) and fructose, i. e. the major components in honey, demonstrated the heaviest loadings both in PC1 and PC2. But it was obvious that this unsupervised model is unable to discriminate the adulterated and true honey. A better statistical method or more emphasis on those minor components is required in order to identify the adulterated honeys from the true ones.

Therefore, a supervised method, LDA, was performed to discriminate the adulterated honey from the true ones. Because of high collinearity in NMR spectral data, PCA was used as a data pretreatment, and the resulting PCs are orthogonal and, of course, independent of each other (Naes & Mevik, 2001). A few principal components with large eigenvalues from PCA model with UV scaling which could describe a complex data set, instead of the raw data, were used as input variables of LDA model.

The validation of the LDA model (applied to the PCA scores) was evaluated by external validation as mentioned above. The first 20 PCs, explaining 84.5% of the total variance of training set, were used to

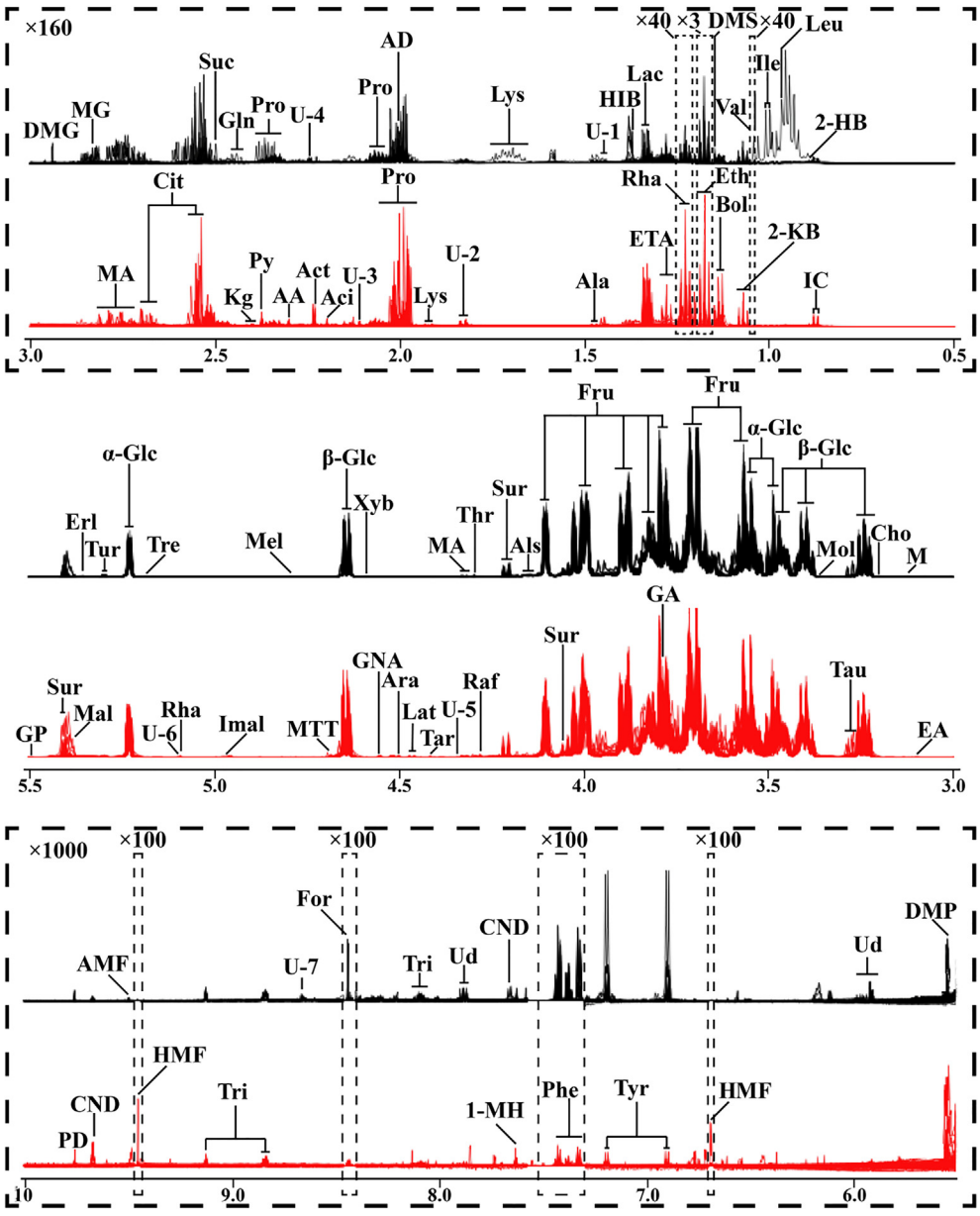

Fig. 1. 600 MHz ^{1}H NMR spectral comparison of true (the black superimposed spectra) and adulterated (the red superimposed spectra) honey. For clarity, the spectral region of 0.5–3.0 ppm was magnified 160 times relative to the spectral region of 3.0–5.5 ppm, among which the region of 1.15–1.20 ppm was magnified 3 times, and the regions of 1.11–1.15 ppm and 1.20–1.25 ppm were magnified 40 times. And the spectral region of 5.5–10.0 ppm was magnified 1000 times relative to the spectral region of 3.0–5.5 ppm, among which the regions of 6.67–6.70 ppm, 7.53–7.56 ppm and 9.44–9.47 ppm were magnified 100 times. The keys of components and spectral information are demonstrated in Table 1.

build the LDA model. Then, the data of train set and test set were respectively used to detect performance of the model. As can be seen in Table 2, the correct classification rate of true honey and adulterated honey were 98.3% and 94.0% in training set, respectively, while the correct classification rate of true honey and adulterated honey were 96.7% and 92.0% in test set, respectively, and the overall prediction rate is 95.8%. The results indicated that the PCA/LDA is able to discriminate true and adulterated honey using ^{1}H NMR spectral data.

3.3. Discrimination of honey adulteration by OPLS-DA

OPLS-DA, another supervised multivariate statistical method, can maximize the differences among different samples by filtering out those variations in the X matrix which were not related to Y variations, which would help to differentiate the adulterated honeys from the true ones. In addition, the NMR spectra reveal that the content of ethanol differed greatly among the real honey and the fake honey. Some samples contained higher levels of ethanol, while some samples had almost no ethanol content. Generally, in the honey adulteration process, it is basically not involved in the incorporation of ethanol, and the difference in the content is more due to storage time or other reasons. Therefore, the ethanol signals were exclude to avoid unwanted effects on the results before OPLS-DA.

As mentioned above, the samples (165 samples) were randomly divided into a training set (110 samples) and a test set (55 samples) for OPLS-DA. The training set was used to build a supervised OPLS-DA with unit variance (UV) scaling in order to emphasize the role of the minor components. As shown in the OPLS-DA scores plot (Fig. 3a), the true honey and adulterated honey samples in training set were obviously separated into two clusters, and the corresponding parameters R^2Y (cum) and Q^2 (cum) with 7-fold cross-validation were 0.691 and 0.644, respectively, indicating a good fitness and high predictive ability. The adulterated honey samples are more dispersible than the true honey samples, which might indicate the large diversity in adulterated patterns. It is notable that two adulterated honey samples were misclassified as true honey, probably because of a lower adulteration concentration of these two samples. A random permutations test (n = 300) was performed to validate the goodness of fit and the predictability of OPLS-DA model and identify the possible overfitting. As shown in Fig. 3b, all R^2 and Q^2 values for the randomly permuted

Table 1
The identified components from the NMR spectra of authentic and adulterated honeys and their relative quantitative determination.

Components	Abbr.	Chemical shift ppm (multiplicity)	Relative concentration	
			True honey	Adulterated honey
1-methylhistidine	1-MH	7.62(s[a])[b]	ND	ND[d]
2,2-dimethylsuccinic acid	DMS	1.15(d)	0.01 ± 0.02[c]	0.02 ± 0.04
2,3-butanediol	Bol	1.13(d)	0.06 ± 0.1	0.2 ± 0.3
2,6-dimethoxyphenol	DMP	5.54(d)	0.09 ± 0.1	0.09 ± 0.2
2-hydroxybutyric acid	2-HB	0.89(t)	0.03 ± 0.04	ND
2-hydroxyisobutyric acid	HIB	1.37(s)	0.01 ± 0.02	0.01 ± 0.01
2-ketobutyric acid	2-KB	1.07(t)	0.02 ± 0.06	0.02 ± 0.07
acetamide	AD	1.99(s)	0.2 ± 0.1	0.1 ± 0.1
acetoacetate	AA	2.30(s)	0.02 ± 0.01	0.03 ± 0.02
acetoin	Aci	1.38(d), 2.20(s)	0.07 ± 0.2	0.05 ± 0.04
acetone	Act	2.23(s)	ND	0.01 ± 0.006
acetoxymethylfurfural	AMF	9.50(s)	ND	0.09 ± 0.4
alanine	Ala	1.48(d)	0.05 ± 0.03	0.04 ± 0.02
allose	Als	4.16(m)	17.9 ± 5.1	14.3 ± 6.1
arabinose	Ara	4.50(d)	1.7 ± 0.6	2.0 ± 0.8
choline	Cho	3.19(s)	ND	0.03 ± 0.07
cinnamaldehyde	CND	6.76(m), 7.64(m), 9.66(q)	ND	0.01 ± 0.03
citrate	Cit	2.53(d), 2.70(d)	0.2 ± 0.4	0.3 ± 0.2
dimethylglycine	DMG	2.93(s)	ND	ND
erlose	Erl	5.35(d)	1.6 ± 1.7	2.0 ± 2.5
ethanol	Eth	1.18(t), 3.64(q)	3.4 ± 9.4	3.8 ± 11.5
ethanolamine	EA	3.12(t)	0.4 ± 0.3	0.4 ± 0.2
ethyl acetate	ETA	1.28(t)	0.2 ± 0.1	0.09 ± 0.1
formic acid	For	8.43(s)	0.09 ± 0.2	0.08 ± 0.07
fructose	Fru	3.56(m), 3.70(m), 3.72(m), 3.79(m), 3.82(m), 3.89(dd), 4.00(m), 4.03(m), 4.10(m)	293.0 ± 47.1	276.6 ± 50.1
glutamine	Gln	2.13(m), 2.45(m)	0.06 ± 0.07	0.06 ± 0.05
guanidoacetate	GA	3.78(s)	77.8 ± 15.5	74.2 ± 13.7
gulonolactone	GNA	4.52(s), 4.55(s)	2.1 ± 0.4	2.2 ± 0.5
hydroxymethylfurfural	HMF	4.66(s), 6.68(d), 7.57(d), 9.45(s)	ND	ND
isocaproic acid	IC	0.87(d)	0.01 ± 0.02	ND
isoleucine	Ile	0.93(t), 1.00(d)	ND	ND
isomaltose	Imal	4.96(d)	0.4 ± 0.4	0.2 ± 0.3
lactic acid	Lac	1.33(d)	0.2 ± 0.1	0.3 ± 0.3
lactose	Lat	4.46(d), 5.20(d)	0.4 ± 0.5	0.7 ± 1.0
leucine	Leu	0.96(t)	0.07 ± 0.4	ND
lysine	Lys	1.72(m), 1.91(m)	0.1 ± 0.04	0.2 ± 0.06
malic acid	MA	2.49(dd), 4.30(dd), 2.73(dd)	9.5 ± 4.0	6.3 ± 3.8
malonic acid	M	3.15(s)	ND	0.02 ± 0.2
maltose	Mal	3.94(m), 3.96(m), 5.38(d)	3.4 ± 4.9	5.4 ± 11.6
maltotriose	MTT	4.67(d), 5.36(d)	4.1 ± 1.7	5.7 ± 3.7
melezitose	Mel	4.78(m), 4.99(d)	0.8 ± 0.4	1.2 ± 0.6
methanol	Mol	3.36(s)	0.6 ± 0.2	0.8 ± 0.4
methylguanidine	MG	2.83(s)	0.01 ± 0.04	ND
phenylalanine	Phe	7.32(m), 7.38(m), 7.42(m)	0.5 ± 1.3	0.1 ± 0.3
proline	Pro	2.00(m), 2.06(m), 2.35(m), 3.33(m)	1.0 ± 0.4	0.6 ± 0.4
pyruvaldehyde	PD	2.15(s), 9.75(s)	ND	ND
pyruvate	Py	2.37(s)	0.05 ± 0.04	0.06 ± 0.03
raffinose	Raf	4.23(d), 4.97(d), 5.43(d)	10.2 ± 8.3	10.2 ± 11.6
rhamnose	Rha	1.23(t), 5.09(d)	1.5 ± 0.6	1.2 ± 0.6
succinic acid	Suc	2.50(s)	0.01 ± 0.02	0.03 ± 0.02
sucrose	Sur	3.56(m), 4.05(t), 4.22(d), 5.40(d)	26.1 ± 33.6	31.2 ± 49.5
tartaric acid	Tar	4.41(s)	0.2 ± 0.2	0.4 ± 0.3
taurine	Tau	3.27(t)	34.1 ± 10.2	43.6 ± 10.1
threonine	Thr	4.25(m)	3.3 ± 1.0	2.4 ± 0.9
trehalose	Tre	5.15(d)	0.2 ± 0.4	0.2 ± 0.4
trigonelline	Tri	4.43(s), 8.08(m), 8.83(m), 9.11(s)	0.02 ± 0.03	0.02 ± 0.03
turanose	Tur	5.30(d)	6.8 ± 3.3	3.0 ± 2.7
tyrosine	Tyr	6.89(d), 7.18(d)	0.04 ± 0.1	ND
unknown-1	U-1	1.45(d)		
unknown-2	U-2	1.82(d)		
unknown-3	U-3	2.11(s)		
unknown-4	U-4	2.24(d)		
unknown-5	U-5	4.34(s)		
unknown-6	U-6	5.11(d)		
unknown-7	U-7	8.65(m)		
uridine	Ud	5.90(d), 5.91(d), 7.87(d)	0.04 ± 0.03	0.01 ± 0.02
valine	Val	0.98(d), 1.03(d)	0.02 ± 0.07	ND
xylobiose	Xyb	4.59(d)	0.2 ± 0.3	0.3 ± 0.3
α-D-galactose 1-phosphate	GP	4.18(s), 5.49(d)	0.6 ± 0.2	0.5 ± 0.2
α-glucose	α-Glc	3.41(m), 3.49(t), 3.53(dd), 3.70(m), 3.82(m), 5.23(d)	164.0 ± 22.2	162.6 ± 18.6
α-ketoglutaric acid	Kg	2.40(t), 3.02(t)	0.05 ± 0.04	0.06 ± 0.04

(*continued on next page*)

Table 1 (*continued*)

Components	Abbr.	Chemical shift ppm (multiplicity)	Relative concentration	
			True honey	Adulterated honey
β-glucose	β-Glc	3.24(dd), 3.40(m), 3.46(dd), 3.72(m), 3.89(dd), 4.64(d)	208.4 ± 27.7	214.5 ± 29.9

[a] s, singlet; d, doublet; t, triplet; q, quartet; dd, doublet of doublets; m, multiplet.
[b] Underlined peaks were selected as the characteristic peaks of the metabolites for quantitative analysis.
[c] The values are shown as mean ± standard deviation.
[d] ND, not detectable or below NMR quantification limit. Signal/noise ratio (S/N) = 10 was defined as the cutoff level of quantification.

models (to the left) are lower than those of the original model (to the right), and the regression line of the R^2 and Q^2 points intersects the vertical axis (on the left) at or below zero (the intercept value of R^2 and Q^2 regression lines were 0.167 and −0.235, respectively), which indicated approving predictive ability without overfitting of the model. The model was also further confirmed by $P_{CV\text{-}ANOVA}$ value close to 0 (9.8×10^{-23}).

Furthermore, the test set was used to evaluate the model by out-of-sample validation, which means to use the model based on training set to predict test set in order to get their prediction rate. Fig. 3c showed the predictive results based on the previous model built by training set, where only one true honey sample was misclassified as adulterated honey and no adulterated honey sample was misclassified if the axis of first latent component (t(1)P) = 0 was set as the dividing line of true and false honey. Fig. 3d depicted the predicted results for the y value in the test set, which was consistent with the results shown in Fig. 3c. As shown in Table 2, the accuracy of classification was respectively 96.0% and 100% for true honeys in training and test sets and respectively 98.3% and 96.7% for adulterated honeys in training and test sets, and the overall prediction rate is 97.6% when combing the training and test sets, indicating that this model was robust and applicable for identification of adulterated honey.

To further identify the specific differential compounds that contributed to the intergroup separation, a volcano plot was drawn based on the results from the univariate statistical analysis (fold change and *t*-test) and the multivariate statistical analysis (correlation coefficient and VIP value from OPLS-DA). As shown in Fig. 4, there are about 30 substances above the dash line ($p < 0.05$), thereinto amino acids such as proline, phenylalanine, and dimethylglycine have lower fold change value and higher p-values, indicating that they have a higher content in the true honey than in the adulterated honey, which generally implying the adulterated substance contains almost no amino acids. Hydroxymethylfurfural (HMF) has a higher content in adulterated honey and used to be regarded as an indicator to detect the presence of invert syrup in honey (Singhal, Kulkarni, & Reg, 1997). However, HMF also often arises from heating or storage of honey. The validity as an adulterant indicator has been therefore questioned (Paradkar & Irudayaraj, 2002). In addition, the amount of acetoin is higher in adulterated honey. Although acetoin is a flavor additive naturally existing in wine, honey, cocoa, butter, coffee, strawberry, garnet berry, etc., it can also be synthesized by a variety of chemical methods (Xiao, Liu, Qin, & Xu, 2007). Therefore, the possibility exists that acetoin levels increase as a result of adulteration containing acetoin as a fragrance. Typical markers selected by three criteria: $p < 0.05$, $|r| > 0.500$ and VIP values at the top of 10% were shown in Table 3. Eight components including proline, xylobiose, uridine, β-glucose, melezitose, turanose, lysine and an unknown component were screened out as potential markers for the discrimination between true and adulterated honey.

In order to understand the discriminatory ability of these eight components, a predictive model of OPLS-DA was also established by use of only their NMR data. As demonstrated in Table 2, the accuracy of classification for true and adulterated honeys was 94.4% (85/90) and 94.7% (71/75), respectively, and the overall prediction rate is 94.5%. Although the accuracy rate of this model is a little lower than the model based on all of the components, it is easier and more practical to establish rapid screening method of honey authentication, thus helping to control the Chinese honey market.

4. Conclusion

Our study provides a fast and promising method of quality control and authenticity of Chinese honey. In this preliminary study, we identified and quantified 65 major and minor components in Chinese

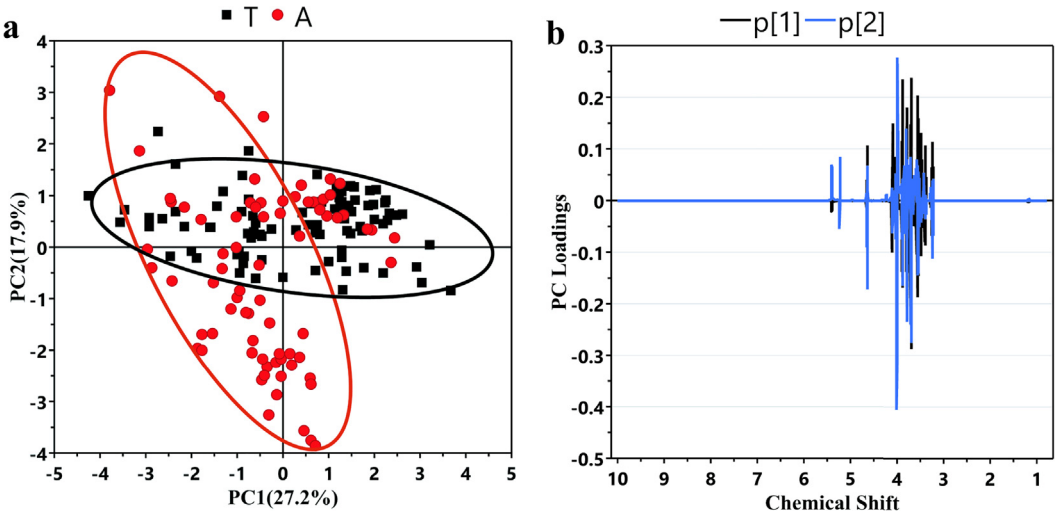

Fig. 2. 2D PCA scores plot (a) and loadings bi-plot (b) (PC1, black line and PC2, blue line) derived from the NMR data of true (T) and adulterated (A) honey samples. The ellipses indicate the respective confidence interval of 95% in the different honey groups.

Table 2
Classification results of honey samples with OPLS-DA and PCA/LDA.

Method	Components in honey	Class	Correct classification rate (%)		Overall accuracy rate
			Training set	Test set	
OPLS-DA	All of the components	T[a]	98.3% (59/60)	96.7% (29/30)	97.6%
		A	96.0% (48/50)	100% (25/25)	
OPLS-DA	The eight discriminating components	T	93.3% (56/60)	96.7% (29/30)	94.5%
		A	98.0% (49/50)	88.0% (22/25)	
PCA/LDA		T	98.3% (59/60)	96.7% (29/30)	95.8%
		A	94.0% (47/50)	92.0% (23/25)	

[a] T, true honey; A, adulterated honey.

honey by NMR spectroscopy. Combing with the multivariate statistical analysis, the NMR data of the honey were utilized to detect honey adulteration and identify the adulteration markers of honey. The accuracy of classification of OPLS-DA for true and adulterated honeys was respectively 97.8% (88/90) and 97.3% (73/75) and the overall prediction rate is 97.6%, while LDA method combined with PCA provided respectively 97.8% and 93.3% accuracy of classification for true and adulterated honeys and an overall 95.8% prediction rate. Moreover, some typical significantly contributing components, such as proline, xylobiose, uridine, β-glucose, melezitose, turanose, lysine and an unknown component, were screened out as the potential markers of discriminating adulterated honey from the true ones. And these markers will help to establish a rapid screening method of honey authentication, and further control the Chinese honey market.

Declaration of Competing Interest

There are no conflicts of interest to declare.

Acknowledgements

This work is financially supported by the National Natural Science

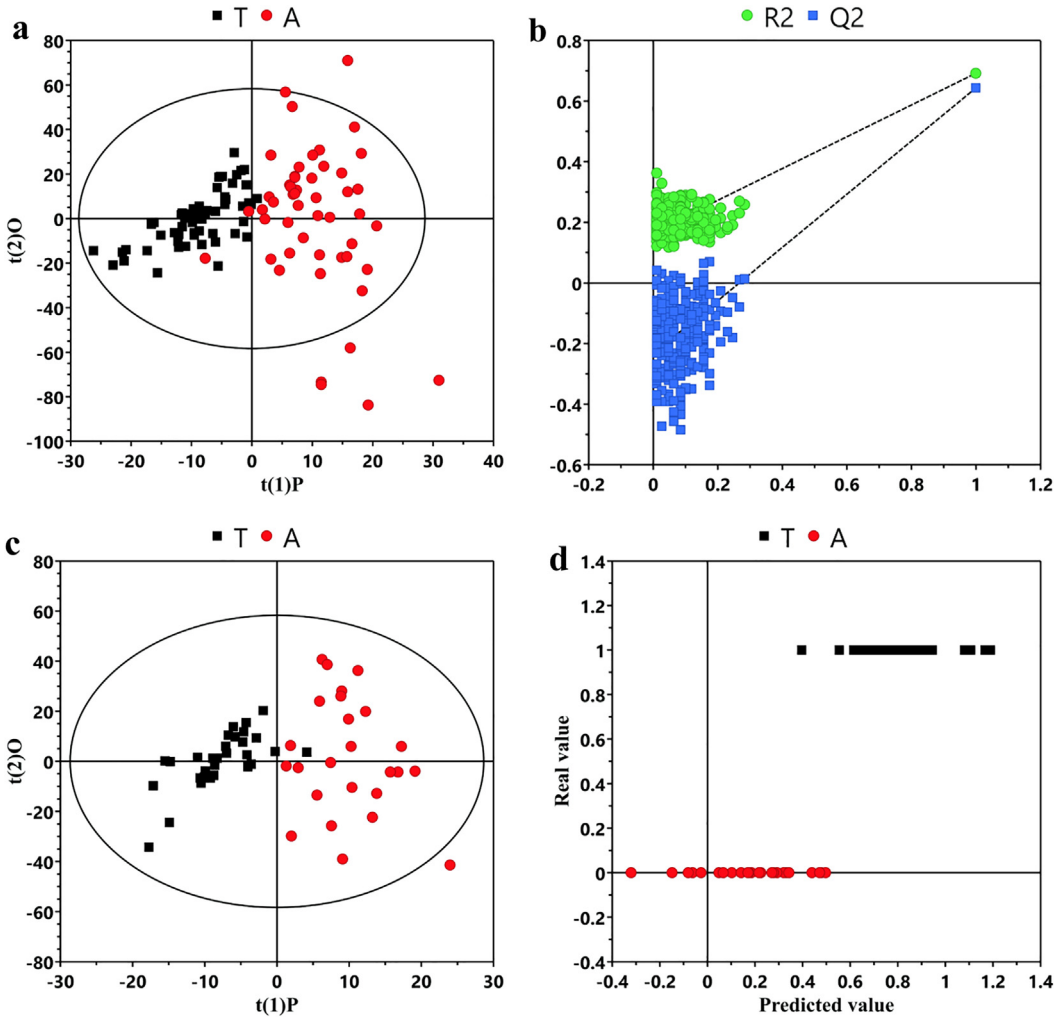

Fig. 3. OPLS-DA models and the validation for the discrimination between the true (T) and adulterated (A) honeys. (a) OPLS-DA discriminative model based on the NMR data of all of the components in training set; (b) the corresponding permutation test (n = 300); (c) OPLS-DA discriminative model based on the NMR data of all of the components in test set; (d) the predicted result in test set.

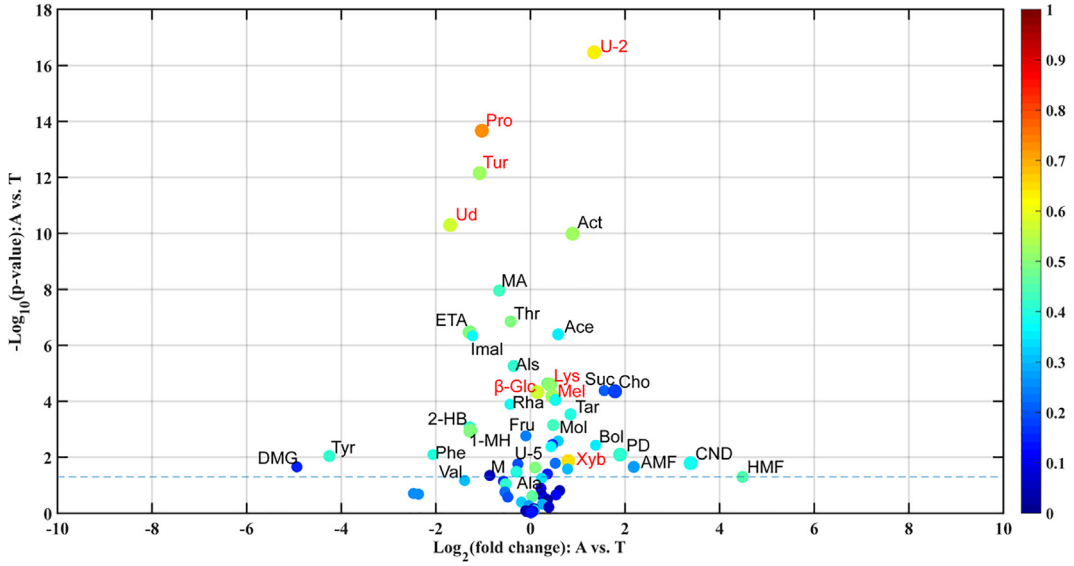

Fig. 4. The volcano plot derived from the univariate (fold change and *t*-test) and multivariate (correlation coefficient and VIP values from OPLS-DA) statistical analysis of the true and adulterated honeys. The keys of the components are shown in Table 1, and detailed information of the screened out components (the metabolites marked by red font) on fold change, p value, correlation coefficient and VIP values are shown in Table 3.

Table 3
The statistical data of the significantly different components between authentic and adulterated honey.

Component	r[a]	p[b]	VIP[c]	FC[d]
Proline	−0.715	2.17×10^{-14}	2.054	0.493
Xylobiose	0.637	0.014	1.824	1.749
Unknown-2	0.623	3.48×10^{-17}	1.780	2.531
Uridine	−0.556	4.97×10^{-11}	1.686	0.310
β-Glucose	0.570	4.54×10^{-5}	1.627	1.105
Melezitose	0.522	6.77×10^{-5}	1.525	1.384
Turanose	−0.523	7.24×10^{-13}	1.497	0.475
Lysine	0.497	2.64×10^{-5}	1.475	1.342

[a] Correlation coefficients, positive and negative signs indicate higher and lower concentration in the adulterated, respectively. The correlation coefficient of $|r| > 0.500$ was used as the cutoff value for the statistical significance based on the discrimination significance at the level of $p = 0.05$.

[b] Variable importance in projection. The top 10% of all VIP values were used as the cutoff values for the statistical significance.

[c] The P-values were transformed by the values Student's *t*-test. The P values < 0.05 were used as the cutoff values for the statistical significance.

[d] Fold change, the concentration ratio between the adulterated and true honeys.

Foundation of China (No. 31671920) and the Natural Science Foundation of Fujian Province (2018Y0078). The authors gratefully acknowledge Mr. Liubin Feng of High-Field NMR Research Center, Xiamen University for technical supports.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2019.108936.

References

Amiry, S., Esmaiili, M., & Alizadeh, M. (2017). Classification of adulterated honeys by multivariate analysis. *Food Chemistry, 224*, 390–397. https://doi.org/10.1016/j.foodchem.2016.12.025.

Bo, Y., Feng, J., Xu, J., Huang, Y., Cai, H., Cui, X., ... Chen, Z. (2019). High-resolution pure shift NMR spectroscopy offers better metabolite discrimination in food quality analysis. *Food Research International, 125*, 108574. https://doi.org/10.1016/j.foodres.2019.108574.

Boffo, E. F., Tavares, L. A., Tobias, A. C. T., Ferreira, M. M. C., & Ferreira, A. G. (2012). Identification of components of Brazilian honey by ^{1}H NMR and classification of its botanical origin by chemometric methods. *LWT – Food Science and Technology, 49*(1), 55–63. https://doi.org/10.1016/j.lwt.2012.04.024.

Consonni, R., & Cagliani, L. R. (2008). Geographical characterization of polyfloral and acacia honeys by nuclear magnetic resonance and chemometrics. *Journal of Agricultural and Food Chemistry, 56*(16), 6873–6880. https://doi.org/10.1021/jf801332r.

Del Campo, G., Zuriarrain, J., Zuriarrain, A., & Berregi, I. (2016). Quantitative determination of carboxylic acids, amino acids, carbohydrates, ethanol and hydroxymethylfurfural in honey by ^{1}H NMR. *Food Chemistry, 196*, 1031–1039. https://doi.org/10.1016/j.foodchem.2015.10.036.

Fang, G., Goh, J. Y., Tay, M., Lau, H. F., Fong, S., & Li, Y. (2013). Characterization of oils and fats by ^{1}H NMR and GC/MS fingerprinting: Classification, prediction and detection of adulteration. *Food Chemistry, 138*(2–3), 1461–1469. https://doi.org/10.1016/j.foodchem.2012.09.136.

Gerhardt, N., Birkenmeier, M., Kuballa, T., Ohmenhaeuser, M., Rohn, S., & Weller, P. (2016). Differentiation of the botanical origin of honeys by fast, non-targeted ^{1}H-NMR profiling and chemometric tools as alternative authenticity screening tool. *Proceedings of the XIII International Conference on the Applications of Magnetic Resonance in Food Science* (pp. 33–37). . https://doi.org/10.1255/mrfs.7.

Kushnir, I. (1979). Sensitive thin layer chromatographic detection of high fructose corn sirup and other adulterants in honey. *Journal – Association of Official Analytical Chemists, 62*(4), 917–920.

Li, M., Li, B., & Zhang, W. (2018). Rapid and non-invasive detection and imaging of the hydrocolloid-injected prawns with low-field NMR and MRI. *Food Chemistry, 242*, 16–21. https://doi.org/10.1016/j.foodchem.2017.08.086.

Li, S., Shan, Y., Zhu, X., Zhang, X., & Ling, G. (2012). Detection of honey adulteration by high fructose corn syrup and maltose syrup using Raman spectroscopy. *Journal of Food Composition and Analysis, 28*(1), 69–74. https://doi.org/10.1016/j.jfca.2012.07.006.

Liu, Y., Qu, F., Luo, L., Xu, W., & Zhong, M. (2019). Detection of rice syrup from acacia honey based on lubrication properties measured by tribology technique. *Tribology International, 129*, 239–245. https://doi.org/10.1016/j.triboint.2018.08.027.

Morales, V., Corzo, N., & Sanz, M. L. (2008). HPAEC-PAD oligosaccharide analysis to detect adulterations of honey with sugar syrups. *Food Chemistry, 107*, 922–928. https://doi.org/10.1016/j.foodchem.2007.08.050.

Musharraf, S. G., Ambreen Fatima, S., Siddiqui, A. J., Iqbal Choudhary, M., & Atta-Ur-Rahman. (2016). 1H-NMR fingerprinting of brown rice syrup as a common adulterant in honey. Analytical Methods, 8(34), 6444–6451. https://doi.org/10.1039/c6ay01082b.

Naes, T., & Mevik, B.-H. (2001). Understanding the collinearity problem in regression and discriminant analysis. *Journal of Chemometrics, 15*(4), 413–426. https://doi.org/10.1002/cem.676.

Ouchemoukh, S., Louaileche, H., & Schweitzer, P. (2007). Physicochemical characteristics and pollen spectrum of some Algerian honeys. *Food Control, 18*(1), 52–58. https://doi.org/10.1016/j.foodcont.2005.08.007.

Paradkar, M., & Irudayaraj, J. (2002). Discrimination and classification of beet and cane inverts in honey by FT-Raman spectroscopy. *Food Chemistry, 76*(2), 231–239. https://doi.org/10.1016/S0308-8146(01)00292-8.

Pontes, M., Marques, J. C., & Câmara, J. S. (2007). Screening of volatile composition from Portuguese multifloral honeys using headspace solid-phase microextraction-gas chromatography–quadrupole mass spectrometry. *Talanta, 74*(1), 91–103. https://doi.org/10.1016/J.TALANTA.2007.05.037.

Rogers, K. M., Sim, M., Stewart, S., Phillips, A., Cooper, J., Douance, C., … Rogers, P. (2014). Investigating C-4 sugar contamination of manuka honey and other New Zealand honey varieties using carbon isotopes. 62(12), 2605–2614. https://doi.org/10.1021/jf404766f.

Ruiz-Matute, A. I., Rodríguez-Sánchez, S., Sanz, M. L., & Martínez-Castro, I. (2010). Detection of adulterations of honey with high fructose syrups from inulin by GC analysis. *Journal of Food Composition and Analysis, 23*(3), 273–276. https://doi.org/10.1016/j.jfca.2009.10.004.

Santos, P. M., Pereira-Filho, E. R., & Colnago, L. A. (2016). Detection and quantification of milk adulteration using time domain nuclear magnetic resonance (TD-NMR). *Microchemical Journal, 124*, 15–19. https://doi.org/10.1016/j.microc.2015.07.013.

Schievano, E., Finotello, C., Uddin, J., Mammi, S., & Piana, L. (2016). Objective definition of monofloral and polyfloral honeys based on NMR metabolomic profiling. *Journal of Agricultural and Food Chemistry, 64*(18), 3645–3652. https://doi.org/10.1021/acs.jafc.6b00619.

Schievano, E., Sbrizza, M., Zuccato, V., Piana, L., & Tessari, M. (2019). NMR carbohydrate profile in tracing acacia honey authenticity. *Food Chemistry, 309*, 125788. https://doi.org/10.1016/j.foodchem.2019.125788.

Shen, G., Huang, Y., Dong, J., Wang, X., Cheng, K. K., Feng, J., … Ye, J. (2018). Metabolic effect of dietary taurine supplementation on Nile tilapia (*Oreochromis niloticus*) evaluated by NMR-based metabolomics. *Journal of Agricultural and Food Chemistry, 66*(1), 368–377. https://doi.org/10.1021/acs.jafc.7b03182.

Singhal, R. S., Kulkarni, P. K., & Reg, D. V. (1997). *Handbook of indices of food quality and authenticity*. Elsevier.

Sobolev, A. P., Thomas, F., Donarski, J., Ingallina, C., Circi, S., Cesare Marincola, F., … Mannina, L. (2019). Use of NMR applications to tackle future food fraud issues. *Trends in Food Science and Technology, 91*, 347–353. https://doi.org/10.1016/j.tifs.2019.07.035.

Sobrino-Gregorio, L., Vargas, M., Chiralt, A., & Escriche, I. (2017). Thermal properties of honey as affected by the addition of sugar syrup. *Journal of Food Engineering, 213*, 69–75. https://doi.org/10.1016/j.jfoodeng.2017.02.014.

Spiteri, M., Rogers, K. M., Jamin, E., Thomas, F., Guyader, S., Lees, M., & Rutledge, D. N. (2017). Combination of ^{1}H NMR and chemometrics to discriminate manuka honey from other floral honey types from Oceania. *Food Chemistry, 217*, 766–772. https://doi.org/10.1016/j.foodchem.2016.09.027.

Venir, E., Spaziani, M., & Maltini, E. (2010). Crystallization in 'Tarassaco' Italian honey studied by DSC. *Food Chemistry, 122*(2), 410–415. https://doi.org/10.1016/j.foodchem.2009.04.012.

Wang, S., Guo, Q., Wang, L., Lin, L., Shi, H., Cao, H., & Cao, B. (2015). Detection of honey adulteration with starch syrup by high performance liquid chromatography. *Food Chemistry, 172*, 669–674. https://doi.org/10.1016/j.foodchem.2014.09.044.

Wu, L., Du, B., Vander Heyden, Y., Chen, L., Zhao, L., Wang, M., & Xue, X. (2017). Recent advancements in detecting sugar-based adulterants in honey – A challenge. *TrAC - Trends in Analytical Chemistry, 86*, 25–38. https://doi.org/10.1016/j.trac.2016.10.013.

Xiao, Z. J., Liu, P. H., Qin, J. Y., & Xu, P. (2007). Statistical optimization of medium components for enhanced acetoin production from molasses and soybean meal hydrolysate. *Applied Microbiology and Biotechnology, 74*(1), 61–68. https://doi.org/10.1007/s00253-006-0646-5.

Zheng, X., Zhao, Y., Wu, H., Dong, J., & Feng, J. (2016). Origin identification and quantitative analysis of honeys by nuclear magnetic resonance and chemometric techniques. *Food Analytical Methods, 9*(6), 1470–1479. https://doi.org/10.1007/s12161-015-0325-1.